DOCUMENT RESUME

ED 450 779 IR 058 035

AUTHOR McGlamery, Patrick
TITLE Issues of Authenticity of Spatial Data.
PUB DATE 2000-08-00
NOTE 8p.; In: IFLA Council and General Conference: Conference Proceedings (66th, Jerusalem, Israel, August 13-18, 2000); see IR 057 981.
AVAILABLE FROM For full text: http://www.ifla.org/IV/ifla66/papers/142-124e.htm.
PUB TYPE Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Cartography; Information Management; *Information Technology; *Maps; *Quality Control; Standards
IDENTIFIERS *Authenticity; Digital Data; Metadata

ABSTRACT

This paper discusses the authenticity of digital spatial data. The first section describes three formats for digital spatial data: vector, raster, and thematic. The second section addresses the integrity of spatial data, including six possible formats for the same information: (1) aerial photographic prints, time stamped, primary, remotely sensed data; (2) orthographic photography, time stamped, primary, remotely sensed data, projected and geographically referenced; (3) cartographic line work, secondary, derived from aerial photography; (4) digital orthographic photography, time stamped, primary, remotely sensed data, projected and geographically referenced; (5) scanned cartography, projected and geographically referenced; and (6) digital cartographic line work, vector digitized from scanned maps or map separates. The third section considers issues of data quality and error, including managing data quality through lineage and metadata. The fourth section covers metadata, including standards for the fields for lineage and process steps. The fifth section summarizes issues of authenticity in spatial data, emphasizing the importance of defining metadata structures to carry document authentication declarations or proofs. (MES)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 450 779

IFLANET Search Contacts
International Federation of Library Associations and Institutions
Annual Conference

IFLA 2000 JERUSALEM
Conference Proceedings

66th IFLA Council and General Conference

Jerusalem, Israel, 13-18 August

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

A.L. Van Wesemael

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

Code Number: 142-124-E
Division Number: II
Professional Group: Special Libraries
Joint Meeting with: -
Meeting Number: 124
Simultaneous Interpretation: *No*

Issues of Authenticity of Spatial Data

Patrick McGlamery
Map and Geographic Information Center
University of Connecticut
Storrs, Connecticut, USA
http://magic.lib.uconn.edu

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Paper

History

Introduction

When we discuss about *"spatial data"* the image that comes to mind is a map. The map as object is, of course, the result of tremendous data compilation. A standard United States Geological Survey (USGS) 7.5 minute topographic quadrangle covers about 140 square kilometers in my part of the United States. These 140 sq. km. are represented on a piece of paper 56 x 68 cm. The map shows rivers, roads, houses and vegetation at a scale of 1:24,000 or 1 cm = 24,000 cm. The data on the map sheets were compiled over decades, both in the field and in a photogrammetric station. The data on a USGS 7.5 min. topo meet national map accuracy standards. These standards state, for example:

> *"Horizontal Accuracy -- For maps on publication scales larger than 1:20,000, not more than 10 percent of the points tested shall be in error by more than 1/30 inch, measured on the publication scale; for maps on publication scales of 1:20,000 or smaller, 1/50 inch. These limits of accuracy shall apply in all cases to positions of well-defined points only. Well-defined points are those that are easily visible or recoverable on the ground, such as the following: monuments or markers, such as bench marks, property boundary monuments; intersections of roads, railroads,*

IR058035

BEST COPY AVAILABLE

etc.; corners of large buildings or structures (or center points of small buildings); etc..."[i]

All USGS maps are printed with, *"Meets National Map Accuracy Standards"* and the seal of the agency. These are clear statements of authenticity, both of the map object and the compilation process. They have served for over half a century to assure the American public of the efficacy of their national mapping program.

Twenty years ago, debate raged over the definition of cartography and maps. The International Cartographic Association (ICA) invited re-definitions of cartography in light of innovations in computer technology. Two camps emerged, stressing the importance of the map on one hand, and the spatial database on the other. M. Visvalingum articulated a middle ground, focusing not on product, but on content. *"If cartography is concerned with the making and use of maps, then it is not just concerned with visual products: it is equally concerned with the processes of mapping, from data collection, transformation and simplification through to symbolism and with map reading, analysis and interpretation. These intellectual processes are expressed in terms of prevailing technologies and computer-based Information Technology is fast becoming the dominant technology of the day."*[ii]

The leap from analog to digital spatial data has been rapid. The inherent nature of spatial data is Cartesian, the points, lines and polygons on maps are imposed on a grid. In fact, the data are collected in the field as numbers in an X/Y, Latitude/Longitude coordinate system. Managing these numbers as digital data is easier than managing them as analog data. Digital cartographic data has been in use since the 1960s, becoming a standard in the past decade with the rapid rise in computing power, the fall in computer prices and education and training of map professionals.

Spatial data

Digital spatial data are generally in three formats; vector, raster and thematic. Vector data are points or nodes linked by arc to represent lines or polygons. This format is most appropriate for line features such as road and hydrographic networks and areal feature like towns, soil types or other geographic areas.

Raster, or image data, is another major format of spatial data. Remotely sensed data such as satellite images, aerial photography, SLAR (Side Looking Airborne Radar) are a major component. Non-image data such as DEMs (Digital Elevation Models), heights above sea level at regularly spaced intervals are also rasters. In addition, paper copies of air photos, maps and plans become raster data when they are scanned. These data become digital spatial objects when they are geo-referenced, linking the rows and columns of numbers to a Cartesian coordinate system. Depending upon the geographic area covered and the resolution of the data, the resulting data files can be very, very large datasets; in the range of 50 Megabytes to several Gigabytes. Color, of course adds to the size and complexity of the data file.

Thematic data that are attributed to features are another level of spatial information. The populations of a town, for example, or its area or tax level are each parts of the digital system. Together vector, raster and thematic spatial data provide powerful tools for analysis and decision making within an system. GIS or Geographic Information Systems have radically effected how we make geographic decisions in the 21st Century.

Integrity of Spatial Data

Determining and assuring the authenticity and integrity of digital spatial information is complex. While spatial information can include cartographic products and attribute data, it always has a geographic component tying it to the earth's surface. Maps, documenting a measurable assurance of spatial accuracy, are fundamental tools of government decision making. In libraries, maps are often the focus of spatial information, and map libraries and collections, either alone or as part of document collections, are often the locus of issues relating to the authenticity of spatial information. In the United States, map libraries often include collections of aerial photography, gazetteer, atlases and guidebooks; and sometimes census information.

Authenticity of printed maps is well established. The United States Geological Survey, the country's primary mapping agency, has legally mandated levels of accuracy and rigid editorial review processes, as have must national mapping agencies. These standards assure accuracy, and integrity. Horizontal, vertical and temporal accuracy are documented, fixed in ink with the information carrier. Spatial information in the United States has given more attention to data quality and spatial accuracy than to authenticity. Anxieties of value and accuracy of the data in the US' litigious society outstripped concerns of data integrity.

Only in the past decade, as the use of spatial data has become ubiquitous in the United States, have issues of authenticity and integrity of data emerged. Spatial data produced by the US federal government are in the public domain. MapQuest, one of the Web's success stories, is based on free federal spatial data. TIGER data, developed by the Geological Survey under contract to the Bureau of the Census created the Topologically Integrated Geographically Encoded and Referenced line feature data at a national scale of 1:100,000. These data began as lines on a map, which were digitized and related to census geographies such as tracts and blocks. In 1990 a free, nationwide set of spatial data were made available to the citizens of the United States. While MapQuest's dataset is generations apart from the '90 TIGER dataset (in fact there have been five subsequent issues of TIGER) they share a common lineage.

One of the complexities of spatial data is that it is not format dependent. The same data can be represented in a variety of ways. For example, here are six graphic formats [not necessarily stages] of the same information.

1. Aerial photographic prints, time stamped, primary, remotely sensed data.
2. Orthographic photography, time stamped, primary, remotely sensed data, projected and geographically referenced.
3. Cartographic line work, secondary, derived from aerial photography.
4. DOQ, digital orthographic photography, time stamped, primary remotely sensed data, projected and geographically referenced.
5. DRG, scanned cartography, projected and geographically referenced.
6. DLG, digital cartographic line work, vector digitized from scanned maps or map separates.

Each of these formats, from aerial photograph, to orthographically mosaiced photograph, to cartographic map, to the digital formats; DLG, DRG and DOQ is a process that looses data. Each step can compromise the integrity of the spatial data.

Data Quality

The overarching issues that have challenged the spatial data community have had to do with data quality and error. Quality is about *"fitness for use."* It has to do with the extent to which a data set, or map output, or a GIS satisfies the needs of the person judging it. Error is the difference between actual data and true data. Error is a major issue in quality. It is often used as an umbrella term to describe all the types of effects that cause data to depart from what they should be. Every GIS action, from conceptualization of the data model to processing of data through to output, has the potential to generate errors and compound existing ones. A user may start with error in one data set (an unreal situation) and through combination with other data sets create an even larger set of errors. The initial error can spread to other data that incorporate the data. The result is information that is less than useful because of the indeterminable compound errors.[iii]

The issues of quality and error can be as mundane as the appropriate scale for the task. For example, census mapping in the United States uses 1:100,000 TIGER data. This scale is appropriate for demographic mapping of the nation. Engineers, however use a finer resolution to build a drainage pipe, 1:1,000. Though one can 'zoom' the census data, the scale of the data continues to be its input scale, not its display scale. If the TIGER data are used to site drainage pipe, it could be off by several tens of meters. The census mapping data are not fit for the use of siting culverts.

Data quality, how to test for it and how to assure it,[iv] has been the topic of several international conferences and workshops. Visualization tools allow the discovery and exploration of error, enabling the user to determine the 'fitness for use'. Quality and error of spatial data can have real and drastic affects. Liability is a subject of great interest and concern in the GIS community. If errors or shortcomings have resulted in inappropriate actions or decisions and parties are harmed, the specter of liability arises for dataset and software producers as well as for other parties involved in the handling of geographic information.[v] Managing data quality, rather than authenticity, has so far been the primary focus of scholarly research.

Managing data quality, to date, has focused on lineage and metadata. A lineage is a record of data history that is presented as a descent or ancestry. The example given above: 1.) aerial photographic prints, 2.) orthographic photography, and 3.) digital orthographic photography move the data from analog to digital, from aspatial artifacts to geo-reference data objects. At each step of the process errors are possible, compromising the integrity and quality of the data. Clearly and effectively communicating and documenting those actions in metadata has emerged as a 'best practices' solution.

Metadata

Describing the process steps in a metadata record is a significant portion of Section 2, Data Quality Information[vi] of the Federal Geographic Data Committee's Content Standards for Digital Geospatial Metadata. The section includes repeatable fields for lineage and process steps. The lineage fields provide the data producer a way to document the source of the data, that is which maps or air photos the data were compiled from. It also gives an opportunity to document process.

> 2.5 Lineage -- information about the events, parameters, and source data which constructed the data set, and information about the responsible

parties.

2.5.1 Source Information -- list of sources and a short discussion of the information contributed by each.

2.5.1.1 Source Citation -- reference for a source data set.

2.5.1.2 Source Scale Denominator -- the denominator of the representative fraction on a map (for example, on a 1:24,000-scale map, the Source Scale Denominator is 24000).

2.5.1.3 Type of Source Media -- the medium of the source data set.

2.5.1.4 Source Time Period of Content -- time period(s) for which the source data set corresponds to the ground.

2.5.1.4.1 Source Currentness Reference -- the basis on which the source time period of content information of the source data set is determined.

2.5.1.5 Source Citation Abbreviation -- short-form alias for the source citation.

2.5.1.6 Source Contribution -- brief statement identifying the information contributed by the source to the data set.

2.5.2 Process Step -- information about a single event.

2.5.2.1 Process Description -- an explanation of the event and related parameters or tolerances.

2.5.2.2 Source Used Citation Abbreviation -- the Source Citation Abbreviation of a data set used in the processing step.

2.5.2.3 Process Date -- the date when the event was completed.

2.5.2.4 Process Time -- the time when the event was completed.

2.5.2.5 Source Produced Citation Abbreviation -- the Source Citation Abbreviation of an intermediate data set that (1) is significant in the opinion of the data producer, (2) is generated in the processing step, and (3) is used in later processing steps.

2.5.2.6 Process Contact -- the party responsible for the processing step information.

The field of this set that is most informative is the 2.5.2.1 Process Description. In a metadata record of soil mapping these fields look like this:

Lineage:
Process_Step:
Process_Description:
Field procedures for the second order soil survey included plotting of soil boundaries determined by field observation and by interpretation of remotely of sensed data. Boundaries were verified at closely spaced intervals, and the soils in each delineation were identified by traversing and transecting the landscape. The classification and map unit names were progressively reviewed December 1993.

Source_Used_Citation_Abbreviation: None
Process_Date: 1994
Source_Produced_Citation_Abbreviation: CTDEP2
Source_Produced_Citation_Abbreviation: CTDEP3
Source_Produced_Citation_Abbreviation: SCS6
Process_Contact:
Contact_Information:
Contact_Organization_Primary:
Contact_Organization:
U.S. Department of Agriculture, Natural Resources Conservation Service
Contact_Address:
Address_Type: Mailing Address
Address: 16 Professional Park Rd.

City: Storrs
State_or_Province: Connecticut
Postal_Code: 06268-1299
Country: USA

There can be several Process_Steps, each one documenting change and possible compromises to the integrity of the data.

Issues of Authenticity in Spatial Data

There are three quite distinct technical and social strategies for asserting authenticity: public, secret, and functionally dependent. Of these three, the public methods are most appropriate for spatial data and in particular *"defining metadata structures to carry document authentication declarations or proofs."* [vii]

This paper is most concerned with the authenticity of spatial data, not necessarily spatial information. Therefore the authenticity of a digital surrogate of a map is not under consideration. Scanned historical maps, for example, should be considered as scanned images, not as spatial data. These data objects lacks spatiality, that is, they are not geographically reference. Scanned air photos are the same, however if the scanned historical map or the air photo is referenced spatially as a data object it becomes spatial data.

In many instances authentication of digital surrogates is well ahead of spatial data. Because of the attention spent on spatial accuracy, data quality and integrity, little attention seems to have been spent on assuring the authenticity of the data. There are no digital certificates, watermarks or other markers, nor does there seem to be any interest in that direction.

There is one tradition in the mapping science that does mark spatial data. Surveyors and their associates often stamp and mark their data. Spatial designs done in AutoCAD are printed and the paper copies notarized. The paper copy becomes the copy of record. A digital interpreation of this is making the maps available in Adobe PDF format and marking them. The Environmental Data Resources, inc. scan and make historic Sanborn Fire Insurance maps available in pdf format, marking each *"page"* with:

The Sanborn Library, LLC This Sanborn Map™ is a certified copy produced by Environmental Data Resources, Inc. under arrangement with The Sanborn Library, LLC. Information on this Sanborn Map™ is derived from Sanborn field surveys conducted in: Copyright © The Sanborn Library, LLC EDR Research Associate Reproduction in whole or in part of any map of The Sanborn Library, LLC may be prohibited without prior written permission from The Sanborn Library, LLC.[viii]

Finally, there needs to be attention paid to determining the authenticity of spatial data, both vector and raster data. Onsrud's observation on the liability factor will drive the research into this area, as will the increasing numbers of GIS users and their opportunities for error. Perhaps libraries can assert and leadership role in this area, documenting and archiving spatial data.

i. United States. Geological Survey, United States National Map Accuracy Standards http://rockyweb.cr.usgs.gov/nmpstds/acrodocs/nmas/NMAS647.PDF, 1947.

ii. Visvalingam, M., *"Cartography, GIS and Maps in Perspective"* Cartographic Journal, 26 (1), 26 - 32, 1989
iii. UNIGIS, 5: Data Acquisition and Data Quality http://td1.ici.ro/lab2_24/lectie.gis/contents.htm
iv. Gary J. Hunter, New tools for handling spatial data quality: moving from academic concepts to practical reality, URISA Journal http://www.urisa.org/Journal/new_tools_for_handling_spatial_d.htm, 1999.
v. Harlon Onsrud, Liability in the use of geographic information systems and geographic datasets. http://www.spatial.maine.edu/~onsrud/pubs/liability40.pdf, 1999.
vi. Content Standard for Digital Geospatial Metadata, http://www.fgdc.gov/metadata/csdgm/02.html
vii. David Bearman and Jennifer Trant, Authenticity of Digital Resources; Towards a Statement of Requirements in the Research Process, D-Lib Magazine, June 1998, http://www.dlib.org/dlib/june98/06bearman.html
viii. Environmental Data Resources, Sanborn Map Report, 1999, http://www.datasite.com/reports/samples/sanborn.pdf

---

Latest Revision: *July 20, 2000*

Copyright © 1995-2000
International Federation of Library Associations and Institutions
www.ifla.org